Sample Size Estimation in the Proportional Hazards Model
for K-sample or Regression Settings

Sample Size Formula for a Normally Distributed Statistic

Suppose a statistic S is known to be normally distributed with mean ω and variance V/n. A hypothesis test having one-sided type I error $\alpha/2$ might be based on a critical function which rejects H_0: $\omega \le \omega_0$ in favor of alternative hypothesis H_1: $\omega > \omega_0$ when

$$Z = \sqrt{n}\,\frac{S-\omega_0}{\sqrt{V}} > z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2); \quad \text{where} \quad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du\,.$$

The power function for this hypothesis test is then

$$Pwr(\omega) = \Pr\left[\sqrt{n}\,\frac{S-\omega_0}{\sqrt{V}} > z_{1-\alpha/2} \mid \omega\right] = \Pr\left[\sqrt{n}\,\frac{S-\omega}{\sqrt{V}} > z_{1-\alpha/2} - \sqrt{n}\,\frac{\omega-\omega_0}{\sqrt{V}} \mid \omega\right] = 1-\Phi\left(z_{1-\alpha/2} - \sqrt{n}\,\frac{\omega-\omega_0}{\sqrt{V}}\right)$$

This power function can be used to

- compute the power β with which the hypothesis test rejects a specific alternative $\omega_1 > \omega_0$ when the sample size is at some given value of n;
- compute the sample size for which a hypothesis test would have prescribed power β to detect a specific "design" alternative $\omega_1 > \omega_0$; or
- compute the alternative $\omega_1 > \omega_0$ which is rejected with prescribed power β when performing the hypothesis test with some given sample size n.

For instance, when desiring to compute a sample size such that the hypothesis test has power β, we merely want

$$Pwr(\omega_1) = \Pr\left[\sqrt{n}\,\frac{S-\omega_0}{\sqrt{V}} > z_{1-\alpha/2} \mid \omega_1\right] = 1-\Phi\left(z_{1-\alpha/2} - \sqrt{n}\,\frac{\omega_1-\omega_0}{\sqrt{V}}\right) = \beta,$$

which then suggests

$$z_{1-\alpha/2} - \sqrt{n}\,\frac{\omega_1-\omega_0}{\sqrt{V}} = z_{1-\beta} = -z_\beta \qquad \Rightarrow \qquad n = \frac{(z_{1-\alpha/2}+z_\beta)^2 V}{(\omega_1-\omega_0)^2}\,.$$

Another approach to sample size estimation is based on the precision with which some parameter can be estimated. For instance, a 100(1-α)% confidence interval for ω might be computed as

$$\left(S - z_{1-\alpha/2}\sqrt{\frac{V}{n}},\quad S + z_{1-\alpha/2}\sqrt{\frac{V}{n}}\right).$$

If we want the width of the confidence interval to be $\omega_1 - \omega_0$ (so the CI will discriminate between the null and alternative hypotheses), then we use sample size formula

$$\omega_1 - \omega_0 = 2\,z_{1-\alpha/2}\sqrt{\frac{V}{n}} \qquad \Rightarrow \qquad n = \frac{(z_{1-\alpha/2}+z_{1-\alpha/2})^2 V}{(\omega_1-\omega_0)^2},$$

which corresponds to the same sample size formula as derived from the hypothesis test, providing we choose power $\beta = 1 - \alpha/2$ (my religion).

Modified Sample Size Formula in the Presence of a Mean-Variance Relationship

Suppose a statistic S is known to be normally distributed with mean ω and variance $V(\omega)/n$. In this case, the variance of the distribution of S depends upon the mean of that distribution—a "mean-variance relationship". The above formulas need to be modified when the variance of the normally distributed statistic depends on the mean. As a general rule, most statisticians ignore this issue because either 1) the sample size will be such that the variance will not differ by very much over the range of alternatives for which the power is, say, between 1% and 99%, or 2) the sample size computation is based on such crude estimates of the variability of the data that any error due to ignoring the mean-variance relationship is negligible, or 3) both. Nevertheless, for completeness I present the modified formulas here for mean-variance relationships. In these formulas, I presume that the power function is higher at ω_0 than at any $\omega < \omega_0$. I note that there are sample size requirements in order to guarantee that the power curve achieves its maximum over the null hypothesis region at this boundary between the null and alternative hypotheses. This requirement is that for all $\omega < \omega_0$, we must have

$$n > z_{1-\alpha/2}\left(\frac{\sqrt{V(\omega)}-\sqrt{V(\omega_0)}}{\omega_0-\omega}\right).$$

This is clearly satisfied when $V(\omega)$ is an increasing function of ω, because in that case, the numerator is negative.

Suppose a statistic S is known to be normally distributed with mean ω and variance $V(\omega)/n$. A hypothesis test having one-sided type I error $\alpha/2$ might be based on a critical function which rejects H_0: $\omega \le \omega_0$ in favor of alternative hypothesis H_1: $\omega > \omega_0$ when

$$Z = \sqrt{n}\,\frac{S-\omega_0}{\sqrt{V(\omega_0)}} > z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2); \quad \text{where} \quad \Phi(z) = \int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du\ .$$

The power function for this hypothesis test is then

$$Pwr(\omega) = \Pr\left[\sqrt{n}\,\frac{S-\omega_0}{\sqrt{V(\omega_0)}} > z_{1-\alpha/2} \mid \omega\right] = \Pr\left[\sqrt{n}\,\frac{S-\omega}{\sqrt{V(\omega)}} > z_{1-\alpha/2}\sqrt{\frac{V(\omega_0)}{V(\omega)}} - \sqrt{n}\,\frac{\omega-\omega_0}{\sqrt{V(\omega)}} \mid \omega\right]$$

$$= 1-\Phi\left(z_{1-\alpha/2}\sqrt{\frac{V(\omega_0)}{V(\omega)}} - \sqrt{n}\,\frac{\omega-\omega_0}{\sqrt{V(\omega)}}\right).$$

This power function can be used to

- compute the power β with which the hypothesis test rejects a specific alternative $\omega_1 > \omega_0$ when the sample size is at some given value of n;
- compute the sample size for which a hypothesis test would have prescribed power β to detect a specific "design" alternative $\omega_1 > \omega_0$; or
- compute the alternative $\omega_1 > \omega_0$ which is rejected with prescribed power β when performing the hypothesis test with some given sample size n.

For instance, when desiring to compute a sample size such that the hypothesis test has power β, we merely want

$$Pwr(\omega_1) = \Pr\left[\sqrt{n}\,\frac{S-\omega_0}{\sqrt{V(\omega_0)}} > z_{1-\alpha/2} \mid \omega_1 \right] = 1 - \Phi\left(z_{1-\alpha/2}\sqrt{\frac{V(\omega_0)}{V(\omega_1)}} - \sqrt{n}\,\frac{\omega_1-\omega_0}{\sqrt{V(\omega_1)}} \right) = \beta,$$

which then suggests

$$z_{1-\alpha/2}\sqrt{\frac{V(\omega_0)}{V(\omega_1)}} - \sqrt{n}\,\frac{\omega_1-\omega_0}{\sqrt{V(\omega_1)}} = z_{1-\beta} = -z_\beta \qquad \Rightarrow \qquad n = \frac{\left(z_{1-\alpha/2}\sqrt{V(\omega_0)} + z_\beta\sqrt{V(\omega_1)}\right)^2}{(\omega_1-\omega_0)^2}.$$

As before, the choice of power $\beta = 1 - \alpha/2$ (my religion) corresponds exactly to choosing sample size according to the precision with which some parameter can be estimated as judged by a 100(1-α)% confidence interval for ω.

When inverting the above power and/or sample size formulas to find the alternative for which a design has prescribed power, it may be the case that an iterative search is necessary.

General Sample Size Formula for 1-sample, 2-sample, and Regression Settings

The *S+SeqTrial Technical Overview* describes a general sample size formula which can be used in the data analysis models most commonly used in the analysis of clinical trial data. (The notation in this document differs slightly from the notation used in the technical overview.) In these models, we let θ represent the measure of treatment effect, which is most often a contrast (difference or ratio) of some within group summary measure μ computed independently for each treatment arm. Statistical analysis can usually be based on an estimate of the treatment effect θ. Most often, either the estimate of θ or the logarithmic transformation of θ are approximately normally distributed in a fixed sample study (i.e., one without interim analyses). We thus let $\omega = g(\theta)$ be the transformed treatment effect measure which is commonly estimated with an approximately normally distributed estimate. The "link" function g( ) is typically the identity function (so $\omega = \theta$) or the logarithmic transformation (so $\omega = \log(\theta)$).

We thus assume that the estimate of ω is approximately normally distributed with

$$\hat{\omega} \sim N\left(\omega, \frac{V}{n}\right)$$

where V is the (average) variability contributed to the estimate by a single observation, and n is the sample size. In general, V can be a function of the within group summary measures μ, as well as other "nuisance" parameters that are independent of μ. In the rest of this document, we ignore any mean-variance relationship. When implementing these formulas, it will generally be necessary to decide whether to make calculations using the value of V under the null, alternative, or some intermediate hypothesis.

Suppose we are interested in discriminating between a null hypothesis H_0: $\omega \le \omega_0$ and an alternative hypothesis H_1: $\omega \ge \omega_1$ in a hypothesis test having one-sided type I error α/2 and statistical power β. When the above approximate distribution holds, sample size computations are most often effected using

$$n = \frac{\delta_{\alpha\beta}^2 V}{\Delta^2}$$

where $\Delta = \omega_1 - \omega_0$ and $\delta_{\alpha\beta}$ is a "standardized alternative", which in a fixed sample study (i.e., one without interim analyses) is $\delta_{\alpha\beta} = z_{1-\alpha/2} + z_{\beta}$.

The same general formula can be used in a group sequential test, providing the estimate of treatment effect can be viewed as a weighted sum of uncorrelated, approximately normally distributed statistics computed on the groups accrued between analyses. This is often referred to as "independent increment structure", and this holds in a wide variety of common clinical trial settings. In these group sequential settings, the "standardized alternative" must be computed using recursive numerical integration of convolutions of densities. (S+SeqTrial will do this for us.)

Use of the General Formula in Common 1-sample Analysis Models

1. *Testing means of continuous distributions:* $Y_i \sim (\mu, \sigma^2)$, $i = 1, \ldots, n$
   - $\theta = \mu$
   - $\omega = \theta$
   - $V = \sigma^2$

2. *Testing geometric means of continuous distributions:* $\log Y_i \sim (\mu, \sigma^2)$, $i = 1, \ldots, n$
   - $\theta = e^{\mu}$
   - $\omega = \log(\theta)$
   - $V = \sigma^2$

3. *Testing proportions of Bernoulli distributions:* $Y_i \sim B(1,\mu)$, $i = 1, \ldots, n$
   - $\theta = \mu$
   - $\omega = \theta$
   - $V = p(1-p)$

Use of the General Formula in Common 2 Independent Sample Analysis Models

1. *Testing means of continuous distributions:* $Y_{ki} \sim (\mu_k, \sigma_k^2)$, $i = 1, \ldots, m_k$; $k = 0,1$
   - $n = m_1 + m_0$
   - Randomization ratio $r = m_1 / m_0$
   - $\theta = \mu_1 - \mu_0$ (difference of means)
   - $\omega = \theta$
   - $V = (r+1)\left[\sigma_1^2 / r + \sigma_0^2 \right]$

2. *Testing geometric means of continuous distributions:* $\log Y_{ki} \sim (\mu_k, \sigma_k^2)$, $i = 1, \ldots, m_k$; $k = 0,1$
   - $n = m_1 + m_0$
   - Randomization ratio $r = m_1 / m_0$
   - $\theta = \exp(\mu_1) / \exp(\mu_0) = \exp(\mu_1 - \mu_0)$ (ratio of geometric means)
   - $\omega = \log(\theta)$
   - $V = (r+1)\left[\sigma_1^2 / r + \sigma_0^2 \right]$

3. *Testing proportions of Bernoulli distributions:* $Y_{ki} \sim B(1, \mu_k)$, $i = 1, \ldots, m_k$; $k = 0,1$
    a. $n = m_1 + m_0$
    b. Randomization ratio $r = m_1 / m_0$
    c. $\theta = \mu_1 - \mu_0$ (difference of proportions)
    d. $\omega = \theta$
    e. $V = (r+1) [p_1 (1 - p_1) / r + p_0 (1 - p_0)]$ (under the alternative)

4. *Testing odds of Bernoulli distributions:* $Y_{ki} \sim B(1, p_k)$, $i = 1, \ldots, m_k$; $k = 0,1$
    a. $n = m_1 + m_0$
    b. Randomization ratio $r = m_1 / m_0$
    c. Odds $\mu_k = p_k / (1 - p_k)$
    d. $\theta = \mu_1 / \mu_0$ (odds ratio)
    e. $\omega = \log (\theta)$
    f. $V = (r+1) [1 / (r p_1 (1 - p_1)) + 1 / (p_0 (1 - p_0))]$ (under alternative)

5. *Testing hazard ratios in survival distributions:* $Y_{ki} \sim S_k(t)$, $i = 1, \ldots, m_k$; $k = 0,1$
    - n = number of **observed events** in both groups combined
    - Randomization ratio $r = m_1 / m_0$
    - Hazard function $h_k(t) = - d (\log S_k(t))$
    - $\theta = h_1(t) / h_0(t)$ (constant ratio of hazard functions)
    - $\omega = \log (\theta)$
    - $V = (r+1) [1 / r + 1]$ (under the null)

<u>*Use of the General Formula in When Comparing Means with Correlated Observations*</u>

1. *("Repeated Measures"):* Suppose the kth treatment group ($k = 0,1$) has m_k independent subjects, each of whom have J measurements, and subjects in different groups are independent
    - $Y_{kij} \sim (\mu_k, \sigma_k^2)$, $k = 0,1$, $i = 1, \ldots, m_k$; $j = 1, \ldots, J$
    - $corr(Y_{kij}, Y_{k'i'j'}) = \rho$ if $k = k'$, $i = i'$, $j \neq j'$
    - $corr(Y_{kij}, Y_{k'i'j'}) = 1$ if $k = k'$, $i = i'$, $j = j'$
    - $corr(Y_{kij}, Y_{k'i'j'}) = \rho$ if $k \neq k'$ or $i \neq i'$
    - Randomization ratio $r = m_1 / m_0$
    - $\theta = \mu_1 - \mu_0$ (difference of means)
    - $\omega = \theta$
    - $V = (r+1) \{ \sigma_1^2 [1+(J-1)\rho]/(J r) + \sigma_0^2 [1+(J-1)\rho]/J \}$

2. *("Crossover"):* Suppose m independent pairs of subjects are randomized such that one member of each pair is in treatment group 0 and one is in treatment group 1.
    - $Y_{ki} \sim (\mu_k, \sigma_k^2)$, $k = 0,1$, $i = 1, \ldots, m$
    - $corr(Y_{ki}, Y_{k'i'}) = \rho$ if $k \neq k'$, $i = i'$
    - $corr(Y_{ki}, Y_{k'i'}) = 1$ if $k = k'$, $i = i'$
    - $corr(Y_{ki}, Y_{k'i'}) = \rho$ if $i \neq i'$
    - $\theta = \mu_1 - \mu_0$ (difference of means)

- $\omega = \theta$
- $V = \{ \sigma_1^2 + \sigma_0^2 - 2 \rho \sigma_0 \sigma_1 \}$

Use of the General Formula in Common Regression Analysis Models

1. *Linear regression (means):* $(Y_i \mid X_i = x_i) \sim (\beta_0 + \beta_1 x_i , \sigma^2), i = 1, \ldots, n$
   - $\theta = E(Y \mid X = x+1) - E (Y \mid X = x) = \beta_1$ (linear contrast of means)
   - $\omega = \theta$
   - $V = \sigma^2 / Var (x)$

2. *Linear regression on log transformed data (geometric means):* $(\log Y_i \mid X_i = x_i) \sim (\beta_0 + \beta_1 x_i , \sigma^2), i = 1, \ldots, n$
   - $\theta = GM(Y \mid X = x+1) / GM (Y \mid X = x) = \exp (\beta_1)$
   - $\omega = \log (\theta)$
   - $V = \sigma^2 / Var (x)$

3. *Logistic regression (odds):* $Y_{ki} \sim B (1, p_k), i = 1, \ldots, m_k ; k = 0,1$
   a. $n = m_1 + m_0$
   b. Randomization ratio $r = m_1 / m_0$
   c. Odds $\mu_k = p_k / (1 - p_k)$
   d. $\theta = \mu_1 / \mu_0$ (odds ratio)
   e. $\omega = \log (\theta)$
   f. $V \cong 1 /[p (1 - p) Var (x)]$ (using an average value for p)

4. *Proportional hazards regression (hazard ratios):* $Y_i \sim S_i(t) , i = 1, \ldots, n$
   - n = number of **observed events** in both groups combined
   - Hazard $h_i(t \mid X_i = x_i) = - d (\log S_i(t \mid X_i = x_i)) = h_0(t) \exp (\beta_1 x_i)$
   - $\theta = h(t \mid X = x +1) / h_i(t \mid X = x) = \exp (\beta_1)$
   - $\omega = \log (\theta)$
   - $V = 1 / Var (x)$ (under the null)

Sample Size Formula for K-sample Setting

The *S+SeqTrial Technical Overview* also provides a sample size formula appropriate when comparing means or geometric means across K independent samples in a fixed sample (no interim analyses) setting. In this setting, we again use some consider some within group summary measure μ_k computed independently for the kth treatment arm, $k=1,\ldots,K$. The null and alternative hypotheses are classically stated as $H_0: \mu_1 = \mu_2 = \ldots = \mu_K$ and $H_1: \mu_i = \mu_j$ for some i,j. Testing of the hypotheses is generally based on the variance of the within group summary measures. That is, the parameter measuring treatment effect is $\theta = Var((\mu_1, \mu_2, \ldots, \mu_K))$. When all groups have equal summary measures, this variance is 0. When the alternative hypothesis is true, the variance across the group summary measures is nonzero.

The exact formula and code used to compute sample sizes in the K-sample setting is given in the technical overview.

Using S+SeqTrial to Compute Number of Events for Proportional Hazards Models

S+SeqTrial provides explicit functions for the computation of sample sizes in the two sample setting for both fixed sample and group sequential trials using the proportional hazards model. Although no explicit facility is provided for proportional hazards regression with a continuous predictor, examination of the results given above for the geometric mean and hazard ratio regressions reveals a similarity of the formulas. In fact, we merely need to use the geometric mean model with a standard deviation of 1 in order to estimate the number of observed events needed for the proportional hazards model.

This also suggests that when planning to use the K-sample logrank statistic, we can merely use the geometric mean model in order to find the number of events needed to provide desired power. In this case, we can use the command line functions (there is a bug in the dialog) to provide a vector of hazard ratios across the K groups. All hazard ratios should be specified relative to the control group, and it will be necessary to include a hazard ratio of 1 reflecting the comparison of the control group to itself.

Computing the Number of Subjects to Accrue to a Survival Study

The above sample size formulas for proportional hazards models provide the number of events needed, rather than the number of subjects. Several approximate approaches are used to determine the number of subjects to accrue:

1. Assume that subjects are accrued uniformly over, say, *(0,a)*, and that data analysis will occur at time $\tau+a$. Further assume exponential survival distribution (a constant hazard) in each group. We can then derive the probability of a subject having an event by the time of analysis, and by dividing the number of events by that probability, derive the number of subjects to accrue. (see *S+SeqTrial Technical Overview*).
2. Under the same assumptions, use the rate of observed events and the average time of follow-up in a Poisson type model.